VIROLOGY

Volume 251, Number 2, November 25, 1998

Contents

Virus Replication/Gene Expression

Virus Structure and Assembly

VIROLOGY
(ISSN 0042-6822)

Published semimonthly (except three times per month in March and September) by Academic Press, 6277 Sea Harbor Drive, Orlando, FL 32887-4900

1998: Volumes 240–252. Prices:
Print journal only: $2395.00 U.S.A. and Canada; $2874.00 all other countries
Annual cumulative CD-ROM only: $1198.00 U.S.A. and Canada; $1437.00 all other countries
Print journal with annual cumulative CD-ROM: $2874.00 U.S.A. and Canada; $3449.00 all other countries
All prices include postage and handling.

Information concerning personal subscription rates may be obtained by writing to the Publishers. All correspondence and subscription orders should be addressed to the office of the Publishers at 6277 Sea Harbor Drive, Orlando, FL 32887-4900. Send notices of change of address to the office of the Publishers at least 6 to 8 weeks in advance. Please include both old and new addresses. POSTMASTER: Send changes of address to V. Drive, Orlando, FL 32887-4900.

Periodicals postage paid at Orlando, FL, and at additional mailing offices.

VIROLOGY 251, 414–426 (1998)
ARTICLE NO. VY989418

Nucleotide Sequence and Transcription Map of Porcine Adenovirus Type 3

P. Seshidhar Reddy,*,[1] Neeraja Idamakanti,* Jae-Young Song,† Joong-Bok Lee,‡ Bang-Hun Hyun,† Jong-Hyeon Park,† Sang-Ho Cha,† Yong-Tae Bae,† Suresh Kumar Tikoo,* and Lorne A. Babiuk*

*Veterinary Infectious Disease Organization, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5E3; †National Veterinary Research Institute, Anyang-si, Kyunggi-do, Korea; and ‡Laboratory of Veterinary Microbiology, Kon-Kuk University, 93-1 Mojingdong, Kwanjingu, Seoul, Korea

Received July 22, 1998; returned to author for revision August 6, 1998; accepted September 10, 1998

The complete nucleotide sequence of porcine adenovirus type 3 was determined and a transcriptional map for the genome was constructed. The size of the genome is 34094 bp in length with an unusually high G + C content (63.7%), the highest thus far reported for any adenovirus. Overall organization of the genome is similar to that for previously sequenced adenoviral DNAs, but there also were distinct differences. The late regions genes are organized into six families, instead of five as they are in human adenovirus type 2. In contrast to bovine adenovirus type 3 and ovine adenovirus, which lack virion-associated RNA genes, the nucleotide sequence analysis of the viral genome indicates that it encodes one short VA RNA species. With the exception of the fiber and a 33-kDa nonstructural protein, the predicted amino acid sequences of the open reading frames in the late regions and the E2 region and IVa2 exhibited a high level of homology, whereas the deduced amino acid sequences of ORFs in E1, E3, and E4 regions, and the pIX showed a lesser homology with the corresponding proteins of other adenoviruses. The proteins V, VII, and IX are unusually long, and the protein VII lacks the consensus protease cleavage site. Genomic and cDNA sequence analysis has identified promoters, cap sites, intron–exon boundaries, polyadenylation signals, and polyadenylation sites in the viral genome.

INTRODUCTION

Adenoviruses constitute a large group of DNA viruses that infect a wide variety of mammals and birds. The group is at present divided into two genera; adenoviruses that infect mammals and share a cross-reacting antigenic determinant on the hexon are grouped under the genus *Mastadenovirus*, whereas those infecting avian species are included in the genus *Aviadenovirus*. Each genus is at present arbitrarily subdivided according to the natural hosts, and the subgroups are further divided into serotypes, which are distinct antigenic entities. Five serotypes have been identified in pigs: porcine adenovirus (PAV) types 1–5 (Derbyshire *et al.*, 1975; Hirahara *et al.*, 1990). Of five serotypes, the type 3 (PAV-3) replicates to the highest titers in cell culture (Hirahara *et al.*, 1990). The prototype strain (6618) of this virus was first isolated from a rectal swab collected from a healthy piglet (Clarke *et al.*, 1967), and experimental infections of piglets with PAV-3 have been subclinical or associated with transient diarrhea (Derbyshire *et al.*, 1975). The proposed use of PAV-3 as a viral vector has stimulated interest in the molecular genetics of the virus. The physical mapping of the genome and the nucleotide sequence analysis of the early region 3 (E3), pVIII, fiber, E4 regions, and the inverted terminal repeats (ITRs) were described recently (Reddy *et al.*, 1993, 1995a, 1995b, 1996, 1997). The nucleotide sequences of the genes encoding the penton base, hexon, protease, and 100 kDa were also reported (McCoy *et al.*, 1996a, 1996b).

Recombinant animal adenovirus-based vectors are promising alternative gene delivery systems in that human adenovirus (HAV)-based vectors have several disadvantages associated with their use (Paillard, 1997). The currently used HAV-based vectors are endemic in most populations, which provide an opportunity for the replication defective viral vectors to multiply either due to complementation or recombination. Preexisting immunity against HAVs also interferes with the entry and long-term expression of a transgene. Moreover, the host range of HAVs is broad and could overcome species barrier without much difficulty, which puts these vectors in a disadvantageous position for regulatory purposes. To circumvent these problems, we are proposing highly species-specific bovine and porcine adenoviruses as gene transfer vectors for vaccine purposes in animals and gene therapy in humans. Recently, the complete nucleotide sequence and transcription map of bovine adenovirus type 3 (BAV-3) was published (Reddy *et al.*, 1998b). The present paper describes the complete nucleotide sequence, genome organization and transcription map for the genome of PAV-3. Availability of the

Published with the permission of the Director of VIDO as journal series no. 247.

[1] To whom reprint requests should be addressed at VIDO, 120 Veterinary Road, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 5E3. Fax: (306) 966-7478. E-mail: reddys@sask.usask.ca.

0042-6822/98 $25.00

TABLE 1

Summary of Published PAV-3 Sequences

GenBank accession no.	Reference	Sequenced region
AF083132	Current report	1–34094
L43077	Reddy *et al.* (1995)	1–144
U24432	McCoy *et al.* (1996)	13556–16283
U34592		19036–21897
U33016	McCoy *et al.* (1996)	21897–22676
U82628		24056–26572
U10433	Reddy *et al.* (1995)	27089–31148
L43363	Reddy *et al.* (1997)	31064–34094

complete sequence information and a transcription map for the whole genome will facilitate the generation of recombinant vectors.

RESULTS AND DISCUSSION

The nucleotide sequences of part of the PAV-3 genome were published earlier and are given in Table 1. As the complete nucleotide sequence of PAV-3 is now available, the nucleotide numbers given in Tables 1 and 2 represent the real distance from the left end of the genome.

Size and structure of the PAV-3 genome

The virus has a genome of 34094 bp with a base composition of 18.3% A, 17.8% T, 31.4% G, and 32.5% C. Thus the G + C content of PAV-3 is 63.7%, the highest reported for any adenovirus. The G + C content of adenovirus genomes sequenced thus far ranges from 33.6% in ovine adenovirus (OAV) (Vrati *et al.*, 1996b) to 55.2% in HAV-2 and -5 (Chroboczek *et al.*, 1992; Roberts *et al.*, 1986). The size of the PAV-3 genome is in agreement with the size determined earlier using restriction enzyme analysis (Reddy *et al.*, 1993) and is comparable to that reported for the genomes of BAV-3 (Reddy *et al.*, 1998b), HAV-12 (Sprengel *et al.*, 1994), and HAV-40 (Davison *et al.*, 1993) but is 1841 nucleotides shorter than HAV-5 DNA (Chroboczek *et al.*, 1992).

A homology search of the GenBank database was made using BLAST for the deduced amino acid sequence for each open reading frame (ORF) with a coding potential of >70 amino acids. Only those ORFs, which could be identified either by their homologies with published adenovirus proteins or by genomic locations, are described here. The organization of the PAV-3 genome as determined based on the sequence analysis of the genome and cDNA clones is presented in Fig. 1 and Table 2.

TABLE 2

Summary of Transcriptional and Translational Features of the PAV-3 Genome

Region	Gene	Cap site	ATG	Splice donor site(s)	Splice acceptor site(s)	Poly(A)$^+$ signal	Poly(A)$^+$ site(s)
E1A	229R	477	533	N	N	1286	1307
	214R		533	1043	1140	1286	1307
E1B	202R	1386	1461	D	D	4085	4110, 4112
	474R	1386	1829	D	D	4085	4110, 4112
pIX	pIX	3367	3394	N	N	4085	4110, 4112
E2A	DBP	27028c	24041c	26949c, 24714c	24793c, 24051c	22560c	22536c
E2B	pTP	27028c	13638c	24949c, 24714c	24793c, 13772c	4075c	4052c
	Pol[b]	27028c	13638c	24949c, 24714c	24793c, 13772c	4075c	4052c
IVa2	IVa2	5864c	5711c	5699c	5441c	4075c	4052c
E3	N	27473	NA	NA	NA	28765	28793
E4	N	33740c	NA	NA	NA	31189c	31170c
L1	52K	6044	10629	9684	10606	13601	13627
	IIIA	6044	11719	9684	11715	13601	13627
L2	pIII	6044	13662	9684	13662	15698[a]	15735
	pVII	6044	15170	9684	15139	15698[a]	15735
L3	pV	6044	15819	9684	15793	18992	19013
	pX	6044	17783	9684	17776	18992	19013
	pVI	6044	18076	9684	18063	18992	19013
L4	Hexon	6044	19097	9684	19096	22544	22567
	Protease[b]	6044	21934	9684	21931	22544	22544
L5	100k	6044	24056	9684	24056	28765	28793
	33K	6044	26181	9684	26130	28765	29793
	pVIII	6044	27089	9684	26792	28765	28793
L6	Fiber	6044	28939	9684	28910	31143	31164

Note. NA, data not available; N, no splicing observed; D, transcripts with different splicing patterns; c, complementary strand.

[a] TTGTTT is present instead of AATAAA.

[b] The splice acceptor sites were predicted based on consensus splice acceptor sequences.

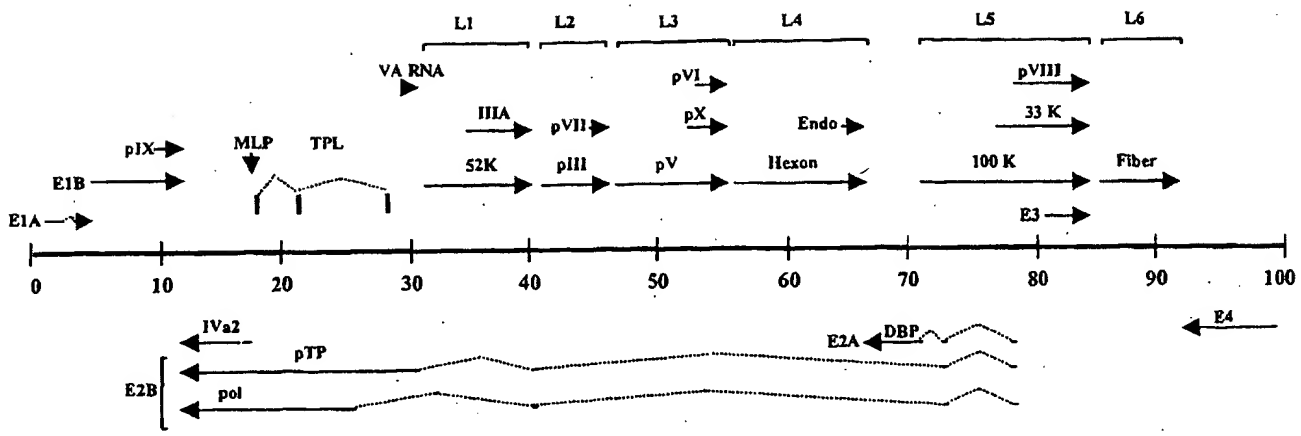

FIG. 1. Genome organization and transcription map of PAV-3. The central solid line, which represents the PAV-3 genome (34,094 bp) is divided into 100 m.u. The genes that are transcribed toward the right are indicated on top and those that are transcribed toward the left are at the bottom of the central line. Solid lines indicate the sequences present in mature mRNA, broken lines represent introns, and arrowheads indicate poly(A)+ sites and show the direction of transcription. The locations of MLP, TPL sequences and VA RNA gene indicated.

Inverted terminal repeats and packaging signal

ITR sequences play an important role in the replication of adenoviral DNA. The ITR sequence of PAV-3 was previously published (Reddy *et al.*, 1995b) and is 144 bp in length. The ITR of PAV-3 could be divided into an AT-rich region and a GC-rich region. The primary sequence of the AT-rich region has a high degree of homology with those of animal adenoviruses, including other PAVs (Reddy *et al.*, 1995b), whereas the GC-rich region of the ITR was not conserved, with only a few short conserved sequence motifs. The packaging domain of adenoviral DNA consists of seven AT-rich repeats and shares a loosely defined consensus motif, 5′-GTN$_{3-4}$TTTG-3′ (Grable and Hearing, 1992). In PAV-3, there are three AT-rich repeats between the left ITR and the E1A promoter, but none of them showed perfect homology with the consensus packaging motifs of HAV-5.

Early regions of the genome

The adenovirus genome is organized into early and late transcription units. From early transcription units, numerous mRNAs are generated, and the encoded polypeptides are required to establish productive viral replication, transformation, and viral latency in infected cells.

E1 region. In PAV-3, the E1A region is located between 1.5 and 3.8 map units (m.u.). Four different types of cDNAs with varying splicing patterns were identified in the E1A region of PAV-3. The pRb binding motif (^{128}Leu-Asp-Cys-Pro-Glu) and a zinc finger motif (^{139}Cys-X$_2$-Cys-X$_{13}$-Cys-X$_2$-Cys) were found in the E1A protein of PAV-3, indicating that the E1A proteins of PAV-3 may have functions similar to those of HAV E1As (Reddy *et al.*, 1998a). The E1B region is located between 4.0 and 12.1 m.u. and shares a poly(A)$^+$ signal and poly(A)$^+$ site with the gene coding for pIX. One E1B mRNA (3 kb) is produced throughout the infection, and five other types of minor mRNAs also are produced late in infection. The ORF analysis of cDNA representing 3.0-kb mRNA showed two overlapping ORFs corresponding to 19- and 55-kDa ORFs of HAV-2 (Reddy *et al.*, 1998a).

E2A and E2B regions. In PAV-3, the mRNAs for DNA-binding protein (DBP), precursor terminal protein (pTP), and DNA polymerase (pol) are transcribed from a promoter located at 79.4 m.u., with a cap site located at nucleotide 27028 (Table 2). The cDNA sequence analysis indicated that the entire coding region of DBP is present in the main body of the mRNA between m.u. 70.5 and 66.5. The two untranslated leader sequences derived from m.u. 79.4 and 75.6 are present in the mRNAs of DBP, pTP, and pol (Fig. 1). The poly(A)$^+$ site of DBP mRNA is at nucleotide 22536, whereas the poly(A)$^+$ signal is found between nucleotides 22560 and 22555. Because we were not successful in obtaining the full-length cDNA representing the pol gene of PAV-3 from the library, the transcriptional data pertaining to pol was derived using the consensus sequences for the pol transcript of HAV-2 (Shu *et al.*, 1988). The transcripts for pTP and pol share a third leader located at 40.4 m.u., which contains the codons for the first three amino acids (Met-Ala-Val) that are common for these two proteins. This exon is located two nucleotides downstream of the poly(A)$^+$ addition site for L1 transcription unit and partly overlaps the ORF of penton base protein. These short exons are spliced to the main body of ORFs of pTP and pol at 30.6 and 25.5 m.u., respectively (Fig. 1 and Table 2). In HAV-2, the initiating Met-Ala-Leu tripeptide, which is common for pTP and pol, lies at m.u. 39. This exon was shown to be important for expression of functional pTP and pol (Shu *et al.*, 1988). The mRNAs of these two proteins also share

TABLE 3

Protein Sequence Identities[a] between PAV-3 and Other Adenoviruses

PAV-3	HAV-2	HAV-12	HAV-40	BAV-3	CAV-1	OAV	CELO
DBP	39.4	41.1	41.8	58.1	46.5	23.8	21.8
pTP	61.3	61.2	61.6	56.3	62.9	25.7	34.4
Pol	46.9	46.4	47.9	46.1	47.1	31.4	31.2
IVa2	65.3	60.6	61.2	62.8	56.5	30.7	25.1
pIX	23.6	25.0	31.8	16.8	28.2	N	N
52K	57.0	53.4	53.7	53.8	56.7	14.9	17.9
IIIA	46.2	49.3	48.3	51.8	52.8	28.2	28.0
III	68.4	68.2	68.4	68.9	71.1	55.5	46.5
pVII	35.1	N	33.9	42.7	39.4	31.5	34.0
pV	37.1	38.3	33	31.5	30.6	N	N
pX	55.6	51.4	54.3	59.7	55.9	31.0	45.8
pVI	34.0	38.1	38.6	26.6	35.3	14.9	21.5
Hexon	67.4	70.2	70.1	68.8	74.0	50.3	47.6
Protease	57.4	53.9	55.9	57.4	56.4	29.4	44.6
100K	49.2	49.4	49.7	48.2	52.5	28.0	29.2
pVIII	61.4	61.0	63.7	57.9	63.7	26.7	17.5
Fiber	25.9	27.5	20.8	28.1	28.6	21.4	21.4

Note. N, homolog not found.

[a] The percentage of identity were determined by using the PALIGN program with default parameters (open gap cost and unit gap cost = 10).

a poly(A)$^+$ signal located between the nucleotides 4075 and 4070 and a poly(A)$^+$ site situated at the nucleotide 4052 with that of IVa2 (Fig. 1 and Table 2). A similar genomic organization has been reported for the E2 regions of HAV-2 (Shu *et al.*, 1988) and BAV-3 (Reddy *et al.*, 1998b).

DBP is a multifunctional protein that contains a variable and highly phosphorylated N-terminal and a highly conserved, nonphosphorylated C-terminal domain (Linne and Philipson, 1980). The variable N-terminal domain is dispensable for viral DNA replication, and the differences noticed in the length of DBPs from various adenovirus serotypes were attributed to the deletions in the N-terminal domains. The N-terminal domain contains the nuclear localization signal (NLS) and is involved in determining the host range (Klessig and Grodzicker, 1979; Morin *et al.*, 1989). The DBP of PAV-3 is 457 amino acids long and contains a poorly conserved N-terminal domain and highly conserved C-terminal domain. The DBP of PAV-3 shows amino acid identities of 21.8–58.1% with those of other adenoviruses (Table 3). The N-terminal domain contains a bipartite NLS (^{47}RRKR and ^{77}RRK). A similar bipartite NLS was found in the DBP of HAV-5 (Morin *et al.*, 1989a, 1989b) and BAV-3 (Reddy *et al.*, 1998b). The DBP of HAV-5 is a phosphoprotein, and the serine and threonine residues present in the N-terminal domain are phosphorylated (Morin *et al.*, 1989a, 1989b). Similarly, the DBP of PAV-3 has 12 serine and threonine residues together in the first 100 amino acids at the N-terminus, which may serve as phosphorylation sites. The C-terminal domain of DBP is involved in DNA binding, initiation and elongation phases of DNA replication, and transcriptional control of major late promoter (MLP). The DBP of HAV-2 contains two zinc atoms in different coordinations (Tucker *et al.*, 1994). The zinc atoms have a structural role and are required for functional activity of the DBP. Structural studies have shown that the cysteine residues located at positions 396, 398, 450, and 467 in the DBP of HAV-2 are involved in binding of one zinc atom (Tucker *et al.*, 1994). All of these cysteine residues (residues 327, 329, 381, and 397), including the distances between them, are conserved in the DBP of PAV-3 (Fig. 2). The second zinc atom in the DBP of HAV-2 is bound by cysteine residues 284, 339, and 355 and one histidine residue at position 286. All of these residues (cysteine residues 215, 270, and 286 and histidine residue 217) and the distances between them are perfectly conserved in the DBP of PAV-3 (Fig. 2). Similar observations were made for the DBPs in HAVs (Kitchingman, 1985), BAV-3 (Reddy *et al.*, 1998b), OAV (Vrati *et al.*, 1996a), and mouse adenovirus type 1 (Cauthen and Spindler, 1996). Two of the 10 conserved glycine residues (280 and 362) in the DBP of HAV-5 were previously shown to be involved in tight β-turns, and three of the remainder (267, 287, and 356) are adjacent to amino acids involved in binding zinc (Tucker *et al.*, 1994). All of these glycine residues (G212, G293 involved in β-turns, and G200, G218, and G287 adjacent to zinc binding residues) are also conserved in the DBP of PAV-3 (Fig. 2). However, only the last two of the five glycine residues (G256 and G340) adjacent to positively charged residues are conserved. These glycine residues in the DBP of HAV-2 were postulated to be involved in interacting with DNA. The C-terminus (residues 513–529) of DBP in HAV-5 forms a hook that sits on a hydrophobic pocket in a second molecule, thereby forming a protein chain. Deletion of the hook destroys

```
P3    MSR-----RYSDVSDYDSGEEGVLIVDDAPVATPSR---RGGGRKRAASPPVQERRKRARA - 53
B3    MNRSHDLTRATSSAD-SSGDENPLVIDETPRKKVSR---KR--RAPIAEPPAEDVTAVKRA - 55
H2    MASREEEQRETTPERGRGAARRPPTMEDVSSPSPSPPPPRAPPKKRLRRRLESEDEEDSSQ - 61
      *..     * .. .  ...    ..... . *.   .  . .  .  .     ..

P3    KPLVSEEEEDEAEETAAATARPARRKGE----------------DGEEEEG---------A - 89
B3    KIFKPSAPAAEAAAPAL-------------------------------------------- - 72
H2    DALVPRTPSPRPSTSTADLAIASKKKKKRPSPKPERPPSPEVIVDSEEEREDVALQMVGFS -122
      . . . . .  .....

P3    RPPRPVTPSPAVARPLLQLDAQEQ---------------------------------KWQR -117
B3    ---------PAASRAL--LEADEV---------------------------------AWQR - 89
H2    NPPVLIKHGKGGKRTVRRLNEDDPVARGMRTQEEKEESSEAESESTVINPLSLPIVSAWEK -183
                . *..  *....                                    * .

P3    AMDLAVQMLVPLKVDVK---GLTLLPDGSTLECFRRGAQAWLNERKISCQLTFSTQKSLLT -175
B3    AMELAVGLCVPLKVDIK---NLTLLPDTGTLECFRKVAQAWLNESKVYLPLTFSTQKTVLT -147
H2    GMEAARALMDKYHVDNDLKANFKLLPD--QVEALAAVCKTWLNEEHRGLQLTFTSNKTFVT -242
      .*. *  .    .** .   ...***    .*.     ..**** .   .***...*. .*

P3    VMARFLLDFVVKAAGLKTPEWNPCGCVIWQHR---SGSEGLHCLHGLPMINKEQLVEMDLN -233
B3    IMGRFLYDFVLKVAGLGTTSWSPTGCVVWRHQCTEGGDGALHCLHGLPMLTKDQVIEMDVN -208
H2    MMGRFLQAYLQSFAEVTYKHHEPTGCALWLHRCAEI-EGELKCLHGSIMINKEHVIEMDVT -302
      .*.*** ... . *..     .* **..* *.     ...*.****  *..*....****.

P3    SENGQRALRETPERAKITTNRWGRNVVQLRNDGAMCCSHDVGSAPNTFSARSCGLFYSEGS -294
B3    SENGQRALKETPHKTKITTNRWNMNVVQLKNEDAACCHFDAALPAGSFTSKSCGMFFSEGP -269
H2    SENGQRALKEQSSKAKIVKNRWGRNVVQISNTDARCCVHDAACPANQFSGKSCGMFFSEGA -363
      ********.* . ..**..***..****..*..* **  *.. ... *...***.*.***.

P3    KAQQAFEQLKAFQKACYPRMGNAETHLLMPLVCDCGWGCRQVPLLGRQTCKITPFALSGSA -355
B3    KALQAFWQIMTFQKACYPRMQSAGTHLLMPLKCDCNWGHSQLPLLGRQVCKITPFNINAGS -330
H2    KAQVAFKQIKAFMQALYPNAQTGHGHLLMPLRCECNSKPGHAPFLGRQLPKLTPFALSNAE -424
      **  ** *...* .* **.  .. .****** *.*.    . *.****  *.***......

P3    ALNPQLVEDPKILASVTHPAVLVFQCYNPVYRGSRGNPQ-KNCDFKISAPDVMLALQLTKQ -415
B3    AVDKSLVEDPKLLASVEHPSVLVFQCCNPVYRQTRANAQ-RNCDFKISAPDMISALQLVKQ -390
H2    DLDADLISDKSVLASVHHPALIVFQCCNPVYRNSRAQGGGPNCDFKISAPDLLNALVMVRS -485
      ...  *..* ..**** **...****.***** .*..    **********.. ** ...

P3    MWAALMQTRPPLTVPEFKWGPQFQVQNTIPPVGTEDDDESLF  -457
B3    MWSALVKQSPPITIPEFRWDPQYQHQNVILPIDQYDADDSLF  -432
H2    LWSENPTELPRMVVPEFKWSTKHQYRNVSLPVAHSDARQNPFDF -529
      .*..     *....***.*... * .*. .*..  *. .. *
```

FIG. 2. Multiple-sequence alignments of DBP homologues. The amino acid sequences of PAV-3 (P3), BAV-3 (B3) and HAV-2 (H2) DBPs were aligned using CLUSTAL of PC GENE program employing default parameters (open gap cost and unit gap cost = 10). Identical and conserved residues are indicated by asterisks and dots, respectively. Dashes indicate gaps. Conserved cysteine and histidine residues that are involved in metal biding are shown in bold and underlined. The NLS are italicized and underlined.

the cooperativity in single stranded (ss)DNA binding, which contributes to the affinity of DBP for ssDNA (Tucker *et al.*, 1994). Some of the highly conserved residues involved in this interaction in the DBP of HAV-5 are L391, L480, V499, and F527; the corresponding residues L322, L410, V429, and F457 were found in the DBP of PAV-3 (Fig. 2).

The pTP of PAV-3 is 631 amino acids in length and is similar to the size reported for HAV-2 and -40 (Shu *et al.*, 1988), BAV-3 (Baxi *et al.*, 1998), and OAV (Vrati *et al.*, 1996a). The pTP of PAV-3 has amino acid identities of 25.7–62.9% to the pTP of other adenoviruses (Table 3). The sequence motif YSRLRYT, which was reported to be conserved among all DNA-terminal proteins studied so far (Hsieh *et al.*, 1990), is also perfectly conserved in the pTP of PAV-3. This motif is believed to play a role in protein primed initiation of DNA replication. The pTP and pol form a stable heterodimer and enter the nucleus together using the nuclear localization signal for pTP (Zhao and Padmanabhan, 1988). The nuclear localization signal RLPV$(R)_6$VP described for HAV-2 and -5 is partly retained in pTP of PAV-3 in the form of ^{345}RLPL$(R)_4$PRP. It has been demonstrated that the pTP of HAV-2 is processed to TP via an intermediate form by the action of viral protease. The sites of cleavage (A–C) of HAV-2 pTP have been determined (Webster *et al.*, 1994). Cleavage at site A generates iTP. The equivalent site is conserved in all HAVs as well as PAV-3. Site B, which is only eight residues C-terminal to site A, is present in HAV-2 and -5 but is absent in PAV-3 and HAV-4. Cleavage at site C in HAV-2 generates mature TP. Site C is conserved in PAV-3 but not in OAV (Vrati *et al.*, 1996). After the assembly of preinitiation complex at the origin of replication, pol catalyzes the transfer of dCMP to serine 580 of the intact

terminal protein (Smart and Stillman, 1982), which acts as primer for initiation of DNA replication. The serine residue and the sequences surrounding it are highly conserved in the pTP of PAV-3 (^{555}NSGD) as in HAVs.

In PAV-3, the coding region of pol partly (50 codons) overlaps the C-terminus of pTP and the N-terminus of IVa2 by 89 codons in different frames. The pol is 1177 residues long and has amino acid identities ranging from 31.2% to 47.9% with its counterparts from other adenoviruses (Table 3). Six regions of homology have been detected among DNA polymerases from diverse eukaryotic and prokaryotic organisms and were thought to be involved in essential functions of this enzyme (Wong *et al.*, 1988). The pol of adenoviruses shares five of six conserved regions with the other members of the α family of pols. Five of these blocks of homology are also present in the primary sequence of PAV-3 pol. The conserved region I (^{987}YGDTDS) is completely retained in PAV-3 pol. In addition to region I, the invariant residues identified in regions II, III, IV, and V of DNA polymerase α family (Wang *et al.*, 1989) are conserved in pol of PAV-3. In pols of adenoviruses, there are two cysteine- and histidine-rich sequences that are conserved between serotypes and might be folded into zinc finger motifs that bind to DNA (Joung and Engler, 1992). In PAV-3 pol, the first cysteine- and histidine-rich sequence is located between amino acids 188 and 212 and is composed of the sequence ^{188}CEHC(X)$_7$HTC(X)$_{10}$HH. The second region is located between the conserved regions I and V near the C-terminus and includes the conserved cysteine residues at 1033, 1037, 1040, 1064, and 1067 [^{1033}CETQCERC(X)$_{23}$CDAC]. The N-terminal region of adenoviral pol contains three clusters of basic sequences (BS), which could function as a nuclear localization signal for the transport of pol (Zhao and Padmanabhan, 1991). Three such motifs, BS I (^{7}RAR), BS II (^{25}RHRPL), and BS III (^{41}RAQRAR), were observed in the pol of PAV-3, which may serve a similar function.

E3 region. The E3 region of PAV-3 is 1179 bp long located between the genes coding for precursor protein VIII and fiber proteins (Reddy *et al.*, 1995a, 1996). It is relatively simple in organization with one poly(A)$^+$ site, which it shares with the L5 region genes. The predicted amino acid sequence of one of the ORFs of the region shows homology with a 13.3-kDa E3 protein of canine adenovirus type 2 (CAV-2). The promoter region of PAV-3 was identified and is located within the body of the gene coding for pVIII (Reddy *et al.*, 1996). Transcriptional mapping of the E3 region of PAV-3 indicated that the L5 region overlaps the E3 region and that the L6 is located immediate downstream of the E3 region. A similar genomic organization was reported for HAV-2; however, there were only five families of late region genes, wherein the L4 region overlaps the E3 region and the L5 region starts just downstream of the E3 region (Cladaras *et al.*, 1985). In BAV-3, the late region genes preceding the E3 region also share poly(A)$^+$ signal and poly(A)$^+$ sites with those of the E3 region (Reddy *et al.*, 1998b). It was suggested that in HAV-5, early in infection, the E3 promoter drives the expression of E3 region, and late in infection, the transcription of the region also initiates from the MLP. Similar observations were made for the E3 region of PAV-3 as one of the cDNA clones representing the E3 region had tripartite leader sequences (TPL) at its 5′ end. Similar observations were made for the E3 region of BAV-3 (Idamakanti, 1998).

E4 region. The E4 transcription unit is located between m.u. 91.4 and 98.9 in the viral genome and is transcribed from the L strand of the genome (Reddy *et al.*, 1997). Of the eight ORFs on the L strand, only the deduced amino acid sequence of ORF 8 showed homology with the 34-kDa E4 protein of HAVs. A partial transcriptional map for the region was also constructed based on the results of the Northern blots, 5′ and 3′ end mapping experiments (Reddy *et al.*, 1997).

Intermediate regions of the genome

Based on the kinetics of expression, the two genes coding for pIX and IVa2 proteins are classified as intermediate region genes, the expression of which begins at about the same time as DNA synthesis (Binger and Flint, 1984). These two proteins are structural components of the virion and were shown to activate the MLP (Lutz *et al.*, 1997; Tribouley, *et al.*, 1994).

pIX. The pIX of HAVs is present in groups of nine, where it strengthens hexon nonamer interactions (Boulanger *et al.*, 1979). In virus-infected cells, this protein acts as a transcriptional activator that is dependent on the presence of TATA box in the promoter (Lutz *et al.*, 1997). In PAV-3, analyses of the nucleotide sequence upstream of ORF representing the pIX showed that the cap site is located at nucleotide 3367; upstream, a TATA box and two GC boxes were found. Like in HAVs, the transcripts of pIX and E1B in PAV-3 are also 3′ coterminal (Fig. 1 and Table 2). The pIX of PAV-3 was 199 amino acids in length and longer than the pIX protein of any other adenovirus for which the nucleotide sequence information is available. Only the N-terminal portion of the protein showed homology with pIX of other adenoviruses. The pIX of PAV-3 is also unusual in that the central portion of the protein has stretches of glutamine (six), alanine (nine), glutamic acid (seven and four), and proline residues (five), the significance of which is not known.

IVa2. In PAV-3, the gene coding for IVa2 is expressed as a spliced mRNA species containing two exons and lacking an intron of 257 nucleotides. The IVa2 and E2B mRNAs share a common poly(A)$^+$ addition site located near 11.8 m.u. (Fig. 1 and Table 2). The transcription start sites of the divergent MLP and IVa2 promoters are only <200 bp apart on the viral genome. The mRNA of IVa2

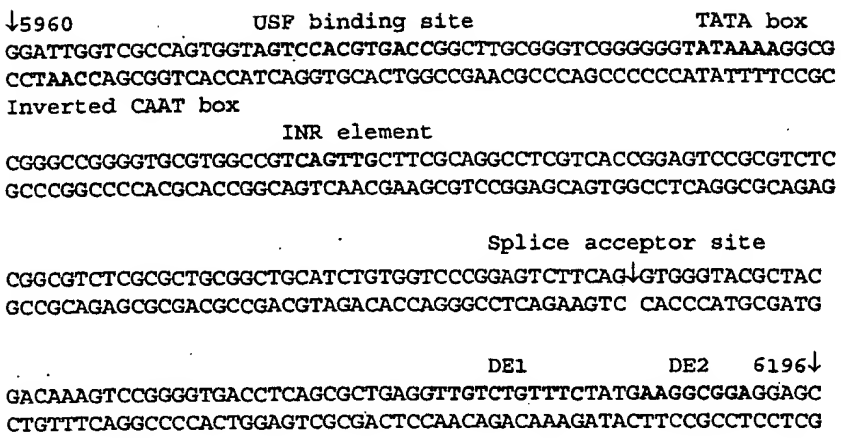

FIG. 3. Sequence showing the transcription control elements of the MLP region in PAV-3. Transcription control elements are shown in bold. The positions of the first and last nucleotides shown in the figure represent the distance from the left end of the genome.

contains a long ORF corresponding to 453 amino acids, of which the first 3 N-terminal amino acids are coded by RNA upstream from the donor splice site and the remaining amino acid residues are coded by RNA downstream from the acceptor splice site. Similar observations were made for the IVa2 transcripts of HAV-2 (Winter and D'Halluin, 1991). The IVa2 is one of the highly conserved structural proteins in PAV-3 and has amino acid identities of 25.1–65.3% to the IVa2 proteins of other adenoviruses (Table 3). The potential nucleoside triphosphate binding site GPTGCGKS, found in IVa2 of adenoviruses (Gorbalenya and Koonin, 1989), is also conserved in IVa2 of PAV-3 between residues 177 and 184, with a serine replacing cysteine at position 181.

Late regions

In HAV-2, transcription initiates predominantly at the MLP, after the initiation of DNA replication, which occurs at ~7 h postinfection (p.i.) (Ziff and Evans, 1978a). Transcription from the MLP accounts for ~30% of the total RNA synthesis at late times. In PAV-3, the DNA replication begins at 14–16 h p.i. (Reddy *et al.*, 1997), after which the majority of transcription is expected to initiate from the MLP. In HAV-2, the MLP is located at 16.8 m.u., whereas in PAV-3, the ML transcription initiates at 17.8 m.u. The MLP of PAV-3 has an inverted CAAT box (5965–5962), the upstream stimulatory factor (USF) binding site (5979–5990), and a canonical TATA box (6009–6015). The initiator element (INR) [TCAGTTG] is located 32 bp downstream from the first T of the TATA box and conforms to the consensus PyPyAN(A/T)PyPy sequence (Javahery *et al.*, 1994) with the exception of the last nucleotide (Fig. 3). In HAV-2, the INR was shown to enhance initiation of transcription from the MLP (Lu *et al.*, 1997). In HAV-2, two sequence elements (DE1, DE2) located downstream of the transcription site were also shown to enhance the transcription from the MLP after the onset of DNA replication. In PAV-3, two such sequence elements are located within the first intron between the nucleotides 6170 and 6180 and the nucleotides 6184 and 6191 (Fig. 3).

In HAV-2, the primary transcript that is initiated from the MLP extends toward the right for ~28 kb and terminates at a position close to 99 m.u. in the viral genome (Fraser *et al.*, 1979). The primary transcripts are cleaved and become polyadenylated at one of the five locations along the genome, generating five families (L1–L5) of mRNAs, each with 3′ cotermini (Ziff and Fraser, 1978b). In PAV-3, there are six families of late mRNAs (L1–L6), each with 3′ cotermini (Fig. 1 and Table 2). In BAV-3, the late regions are organized into seven families (Reddy *et al.*, 1998b). This finding might have important implications in the development of adenoviral vectors as it should be possible to insert a foreign gene under the control of appropriate splicing elements as a new family of late genes and obtain high levels of expression of the gene. In HAV-2, after the selection of a 3′ end, the primary transcript is processed by splicing in such a way that each mature mRNA gains a common set of three short 5′ leader segments called TPL derived from m.u. 16.8, 19.8, and 26.9 (Berget *et al.*, 1977). The TPL enhances the translation of mRNA in infected and transfected cells. Although the INR element was located between the nucleotides 6040 and 6046, the nucleotide sequence determination of several cDNA clones representing late transcripts of PAV-3 had their 5′ ends at the nucleotide 6064. Often the cDNA clones do not contain the complete 5′ end sequences of mRNAs, and the transcription start site of the MLP used in the text is based on an INR element. Thus the first leader of the TPL in PAV-3 lies next to the MLP and is predicted to be 81 nucleotides long. The second leader lies within the gene coding for pol and is 68 nucleotides in length. The MLP and the first and second leaders of the TPL are located within the

coding region of the pol gene. The third leader, the longest of all (99 nucleotides)s is located within the gene coding for pTP. Thus the TPL of PAV-3 is predicted to be 248 nucleotides long and is derived from three exons located at 17.7, 20.9s and 28.1 m.u. In HAV-2, the total length of the TPL is 201 nucleotides, with its first exon of 41 nucleotides from the MLP region, its second exon of 71 nucleotides from DNA pol region, and its third exon of 89 nucleotides from the pTP region of the genome (Chow *et al.*, 1977). Similarly, in BAV-3 the length of the TPL is 205 nucleotides long and the three exons are derived from different parts of the genome located at 16.2, 19.3 and 26.4 m.u. (Reddy *et al.*, 1998b). In OAV, the length of the TPL is only 157 nucleotides, with its first exon of 31 nucleotides from the MLP region and its second and third exons of 63 nucleotides each from the pTP region (Vrati *et al.*, 1996a).

L1 region. The mRNAs of the L1 region are characterized by having a common poly(A)$^+$ site at m.u. 39.9 (Fig. 1 and Table 2). Two mRNAs with splice acceptor sites located at m.u. 31.1 (52 kDa) and 34.4 (pIIIa) have been identified. A similar organization has been reported for the L1 region of HAV-2 (Akusjarvi and Persson, 1981). In HAV-5, the 52/55-kDa proteins represent two differentially phosphorylated forms of a 48-kDa precursor protein that play a major role during the assembly of infectious virus particles (Hasson *et al.*, 1992). The 52-kDa protein of PAV-3 is 335 amino acids in length, and is 80 amino acids shorter than the corresponding protein in HAV-2. Shorter 52-kDa proteins are also noticed in BAV-3 (Reddy *et al.*, 1998b) and OAV (Vrati *et al.*, 1996b), which could be accounted for by deletions in the N- and C-terminal portions of the respective proteins. Nevertheless, the central portion of the PAV-3 52-kDa protein exhibits >50% amino acid identity with those of HAV-2, -12, and -40 and BAV-3 (Table 3).

The protein pIIIa is a structural component of the virion that makes contact with the outer aspect of the hexon. (Stewart *et al.*, 1993). In PAV-3, the IIIa protein is 622 amino acids long, the longest thus far reported for adenoviruses. The pIIIa of PAV-3 is also unusual in that it lacks a consensus protease cleavage sequence, found in the pIIIa of HAVs. The pIIIa protein of MAV-1 also lacks a consensus protease sequence (Meissner *et al.*, 1997); however, pIIIa of PAV-3 has amino acid identities of 28–52.8% to the pIIIa of other adenoviruses (Table 3).

L2 region. The L2 region has a poly(A)$^+$ tail addition site at m.u. 46.1. Two major mRNA species with splice sites at coordinates 40.0 (pIII) and 44.4 (pVII) have been identified (Fig. 1 and Table 2). The penton base protein (pIII) is known to play an important role in virus entry into cells (Wicham *et al.*, 1993). After attachment to cells via the fiber protein, the penton base protein (pIII) of adenoviruses interacts with αv integrins and promotes efficient uptake of virus into cells (Mathias *et al.*, 1994). The pIII of PAV-3 is 484 amino acids in length (McCoy *et al.*, 1996a) and, next to the hexon, shows the highest degree of homology with those of other adenoviruses (Table 3). The PAV-3 protein is similar in size to those of BAV-3 (Reddy *et al.*, 1998b), CAV-1 (Morrison *et al.*, 1997), and MAV-1 (Meissner *et al.*, 1997) but is much shorter due to deletions near the N-terminus and central portions of the protein relative to HAV-2. The penton base proteins of most of HAVs have a conserved Arg-Gly-Asp (RGD) sequence, which is predicted to lie at the apex of two extended α helices. The RGD sequence motif promotes efficient uptake of virus into the cellular endosomes. Such a sequence motif is absent in the penton base protein of PAV-3. The RGD motif is also absent in the penton base proteins of HAV-40 (Davison *et al.*, 1993), BAV-3 (Reddy *et al.*, 1998b), OAV (Vrati *et al.*, 1996b), CAV-1 (Morrison *et al.*, 1997), egg drop syndrome virus 76 (Rohn *et al.*, 1997), and MAV-1 (Meissner *et al.*, 1997). It is possible that these viruses might use non-RGD motifs in entering cells because non-RGD sequences are also capable of mediating integrin binding (Hynes *et al.*, 1992). One such motif is LDV, which is partly conserved in the penton base proteins of PAV-3 as LDL. In HAV-2, the penton base protein interacts with the fiber in the virion through a conserved domain (SRLSNLLGIRKR) which is also found in the pIII of PAV-3 (^{242}SRLNNLLGIRKR). A similar domain was found in the penton base protein of BAV-3 (Reddy *et al.*, 1998b).

The polypeptide pVII is a major protein associated with adenovirus cores. It accounts for 10% of the protein mass of the virion (Sung *et al.*, 1977) and is involved in the packaging of viral DNA. In PAV-3, the pVII is 171 amino acids long and shows a high degree of sequence conservation with other adenoviruses in the N-terminal 50 residues, including the consensus protease cleavage site (^{20}MYGG-ARRL). The remainder of the protein is not well conserved, and the lack of conservation of the C-terminus of pVII has been noted for other animal adenoviruses. The amino acid sequence analysis of PAV-3 pVII indicated that the protein is very rich in basic amino acids (34 of 171 residues), which is consistent with the function of the protein in condensation of viral DNA.

L3 region. The L3 mRNAs have a poly(A)$^+$ tail addition site at 55.8 m.u. Three mRNA species coding pV, pX, and pVI are transcribed from this region (Fig. 1 and Table 2). The protein pV is involved in directing the adenovirus DNA to the nucleus after uncoating. The ORF coding for the PAV-3 pV is 629 amino acids in length, which is the longest pV reported for any adenovirus. This is mainly due to insertion of amino acids in the central portion of the protein relative to other pV proteins. The pV protein of PAV-3 shows homologies with those of other adenoviruses, which range from 31% to 38% (Table 3). The pV of PAV-3 contains ~20% (123 of 629 residues) basic amino acids and two bipartite NLS: one at the N-terminus (^{42}KKRRKAKRX$_{13}$RPRRR) and the other at the C-terminus of the protein (^{586}RWQRRX$_9$RLPRR). Similar observations

were made for the pV of HAV-2 and BAV-3 (Reddy *et al.*, 1998b; Russell and Kemp, 1995). The presence of two NLS possibly increases the rate of nuclear transport of this protein. These properties of the protein are likely to be related to its DNA binding function.

In PAV-3, the precursor for pX has 72 amino acids compared with 80 residues for pX of HAV-2 and BAV-3 (Reddy *et al.*, 1998b; Roberts *et al.*, 1986). Like in HAV-2, the pX of PAV-3 has retained both protease cleavage sites: one near the N-terminus and the other in the middle of the protein. The main difference in length results from a deletion immediately before the amino terminal proteinase cleavage site. The N-terminal proteinase cleavage site does not completely comply with the consensus protease cleavage site in that it has valine in the P3′ position. In HAV-2, the pX is a 19-amino-acid peptide that binds strongly to DNA (Anderson, 1990). Based on the proteinase cleavage sites, the pX of PAV-3 is expected to be 22 residues long. The C-terminal fragments of pX of PAV-3 and HAV-2 show high homology (67% identical amino acids).

The protein pVI of PAV-3 is 292 amino acids in length, 25 residues longer than the HAV-40 pVI (Davison *et al.*, 1993). The difference in lengths is mainly due to two insertions in the central portions of the protein, a least conserved portion of the protein. The two protease cleavage sites, one at the N-terminus (^{30}MNGG-ALSW) and the other near the C-terminus (^{278}IVGL-GVRT), conform to the consensus protease cleavage site sequences. The N-terminal cleavage product is predicted to be 33 residues, the same length as in the HAV-2 N-terminal fragment. The C-terminal fragment is expected to be 12 residues long, one residue longer than that seen in HAV-2 (Anderson, 1990).

L4 region. The L4 mRNAs are characterized by a poly(A)$^+$ tail addition site at 66.2 m.u. Two mRNA species, encoding the hexon and a protease, are transcribed from this region (Fig. 1 and Table 2). The hexon is the most abundant structural protein of adenovirus and forms 240 of the 252 capsomers and contains both group- and type-specific epitopes (Pettersson, 1984). The hexon of PAV-3 is 939 amino acids in length, which is similar in size to that reported for the hexon of BAV-3 (Reddy et al., 1998b). The sequence alignment of PAV-3 hexon with that of HAV-2 show significant homology throughout the molecule, except in three regions of the protein. The first region lies between the residues 133 and 179. and the second one lies between the residues 233 and 258 (numbering with reference to PAV-3). These regions fall in the base domain P1 and 1_1 loop of the hexon. The third region, which did not show homology, is located between the residues 399 and 437 and corresponds to loop 1_2 of the molecule. Type-specific antigenic sites have been demonstrated in loops 1_1 and 1_2 in the hexon proteins of HAV-2 and -5 (Toogood *et al.*, 1992). Although the sequence forming the loops is not well conserved, the sequence elements required to stabilize the structure are well conserved, and the overall structure of the PAV-3 hexon protein would be expected to be similar to that published for HAV-2 and MAV-1 (Weber *et al.*, 1994).

The proteinase of PAV-3 is 204 amino acids in length (McCoy *et al.*, 1996b), similar to those from other adenoviruses, which range between 201 and 214 amino acids (Weber, 1995). Homology as a percentage of identity with other adenoviruses varies between 29.4% and 57.4% (Table 3). In the proteinase of HAV-2, histidine-54 and cysteine-122 were shown as catalytic site residues (Rancourt *et al.*, 1994). These two and other residues, which are involved in catalytic activity, are also conserved in the proteinase of PAV-3. Adenoviral proteinases are basic in nature, and this property may help it to be packaged into virions (Webster *et al.*, 1993). Similarly, there are 27 basic residues in the PAV-3 proteinase sequence. The proteinase of HAV-2 requires a peptide (GVQSLKRRRCF) for its activity, which is derived from the C-terminus of pVI (Mangel *et al.*, 1993; Webster *et al.*, 1993). This peptide is highly conserved in other serotypes (Weber, 1995), including the peptide from the C-terminus (^{280}LGVRTCK-RRRCY) of PAV-3 pVI. Because the PAV-3 proteinase and the C-terminus of pVI share significant homologies with the corresponding proteins from other adenoviruses, the same mechanism of cleavage might be used during virus assembly in PAV-3.

L5 region. L5 region is characterized by a common poly(A)$^+$ tail addition site at 84.4 m.u. It codes for three polypeptides, two of which (100 and 33 kDa) are nonstructural, and the other is pVIII (Fig. 1 and Table 2). The 100-kDa protein of PAV-3 is 838 amino acids long (McCoy *et al.*, 1996b) and shows 28.0–52.5% amino acid identities with 100-kDa proteins of other adenoviruses (Table 3). In HAV-2, the 100-kDa polypeptide is made in large quantities late after infection. It is specifically associated with the hexon in adenovirus infected cells and plays an essential role in trimerization of the hexons (Cepko and Sharp, 1983). The 100-kDa protein of HAV-5 enhances the translational efficiency of late viral mRNAs by binding to them (Riley and Flint, 1993). Sequences that conform to a consensus RNA binding region of RNA binding proteins were found in adenovirus 100-kDa proteins (Hayes *et al.*, 1990). The common RNA recognition sequence consistis of four domains. Sequences that conform to each of the subdomain consensus sequences can also be found in the 100-kDa protein of PAV-3, suggesting the importance of these domains in the protein function. The ORF corresponding to a 33-kDa phosphoprotein of HAV-2 is 225 amino acids in length in PAV-3 and did not show any homology with the corresponding proteins from other adenoviruses.

Polypeptide pVIII, which is one of the hexon-associated proteins, connects the core with the inner surface of the adenovirus capsid. The pVIII of PAV-3 is 223 amino

acids long (Reddy *et al.*, 1995a) and shows the highest degree of homology with those of HAV-40 and CAV-1 (Table 3). The N- and C-terminal regions of the protein show higher homologies than the central region. Two potential protease cleavage sites (^{108}LAGG-GALA and ^{150}LGGG-GRSS) that conform to the consensus cleavage sites were found in the pVIII of PAV-3.

L6 region. In PAV-3, the L6 mRNA is characterized by a poly(A)$^+$ addition site at 91.4 m.u. A single polypeptide fiber is translated from this region (Fig. 1 and Table 2). In HAV-2, ~30% of the mRNA population coding for fiber has been shown to contain ancillary x, y, and z leader sequences in addition to the TPL (Anderson and Klessig, 1984). No such leader sequences were found in the cDNAs representing the fiber mRNA of PAV-3. The fiber of PAV-3 is 447 amino acids long and could be divided into tail, shaft, and knob regions. The shaft component could be arranged into 14 pseudorepeats, and each repeat is 15 residues long (Reddy *et al.*, 1995a). The fiber of PAV-3 shows amino acid identities of 21.4–28.6% to the fiber proteins of other adenoviruses, and the N-terminal portion showed the highest homology (Table 3). In HAV-2, the tail of the fiber protein contains a hydrophobic sequence (FNPVYPYD) and is thought to be involved in interaction between the fiber and the penton base protein (Caillet-Boudin, 1989). A similar sequence motif (^{13}FDPVYPYD) is also present in the tail portion of the fiber in PAV-3. In most of the adenoviruses, a conserved sequence motif separates the head from the shaft region of the fiber (Chroboczek *et al.*, 1995). This sequence motif (^{260}TLWT) is also perfectly conserved in the fiber of PAV-3.

Virion-associated RNA. HAVs encode low-molecular weight RNAs, called virion-associated (VA) RNAs, which are transcribed by RNA polymerase III. These RNAs are required for efficient translation of viral mRNAs late after infection. In HAVs, VA RNA genes are located between the genes coding for pTP and 52/55-kDa proteins. The presence of a 164-bp sequence between the splice acceptor sites of pTP and the 52/55-kDa protein genes and additional signature sequences in the region indicates that there is one short VA RNA gene in the genome of PAV-3 (Fig. 1). Nucleotide sequence analysis of the gene revealed several regions, which showed complementarity to the other regions within the locus. Potential structural models differing in calculated stability for the VA RNA sequence were derived by computer-assisted prediction of folding patterns using the Zuker algorthim (1981). The length of VA RNAs in HAVs is ~160 nucleotides (Ma and Mathews, 1996). Based on the upstream and downstream signature sequences, the VA RNA of PAV-3 is expected to be ~140 bases in length. Shorter VA RNAs were also reported in HAV-10 (Ma and Mathews, 1996) and CELO (Larsson *et al.*, 1986). The smaller size of VA RNA in HAV-10 was attributed to a deletion of 27 nucleotides as a result of looping out between flanking direct repeats (Ma and Mathews, 1996). Upstream of the VA RNAs, 28–33 nucleotides rich in adenosine separate the VA RNA initiation site from pTP splice site. In PAV-3, there are seven adenosine residues upstream of the splice acceptor site of pTP. In HAVs, the run of T residues located at the end of VA RNAs acts as termination signal. There are two runs of T residues separated by adenosine immediately upstream of splice acceptor site of the 52/55-kDa protein, which could act as a similar termination signal for VA RNA in PAV-3.

In summary, the ORFs for all the structural and nonstructural proteins of PAV-3 are located in the same relative positions in the genome as in other HAVs. However, a relatively high G + C content, organization of the late region genes into six families, the absence of additional leader sequences (x, y, and z) in the fiber, and the presence of a single small VA RNA gene are some of the distinctive features of the PAV-3 genome. Other distinctive features of the genome include a smaller and simple E3 region; unusually long pV, pVII, and pIX ORFs; the absence of a protease cleavage site in pVII; and the absence of RGD motif from the penton base.

MATERIALS AND METHODS

Virus and viral DNA

The 6618 strain of PAV-3 was cultivated in a swine testis (ST) cell line. The cells were grown in Eagle's minimum essential medium supplemented with 5% fetal bovine serum. The purification of virus and extraction of DNA from virus were carried out as described previously (Graham and Prevec, 1991).

Plasmids and genomic DNA sequencing. Selected restriction enzyme fragments of PAV-3 DNA were cloned into pGEM-3Z and pGEM-7Zf(+) plasmids (Promega). Nested set deletions were made using exonuclease III and S1 nuclease as described by Henikoff (1984). Nucleotide sequences were determined on both strands of the genome by the dideoxy chain-termination method (Sanger *et al.*, 1977) with Sequenase enzyme (U.S. Biochemicals) and by the dye-terminator method with an Applied Biosystems (Foster) DNA sequencer. Sequence alignments were carried out by using PALIGN and CLUSTAL programs of the PC-GENE sequence analysis software package (Oxford Molecular) and Clone Manager (version 4).

cDNA library. Two cDNA libraries were made using polyadenylated [poly(A)$^+$] RNA extracted from PAV-3 infected ST cells at 12 and 24 h p.i. The cDNAs were synthesized using the reagents from Stratagene and cloned into the Uni-Zap vector. The cDNA library made with RNA extracted at 12 h p.i. was used to screen for the cDNAs of early regions of the genome, and the one made using RNA extracted from 24 h p.i. was used for cDNAs of late region genes. To select the cDNA clones representing the mRNAs of early region of the genome, the

restriction enzyme fragments corresponding to the 5′ ends of the regions were labeled with ^{32}P and used as probes. Similarly for late region genes, ^{32}P-labeled TPL sequence was used as a probe. The plaques that hybridized to specific restriction enzyme fragments of PAV-3 were plaque purified twice, and the pBluescript SK(+) containing cDNAs were excised from Uni-Zap vector according to the Stratagene protocol. The resulting clones were subjected to restriction enzyme analysis and grouped. The nucleotide sequences of the representative cDNA clones from each group were determined.

Nucleotide sequence accession number. The complete nucleotide sequence reported here has been submitted to GenBank and assigned accession no. AF083132.

ACKNOWLEDGMENTS

We thank A. Zakhartchouk, M. K. Baxi, C. Pyne, Yan Chen, and R. Brownlie for their help and suggestions. This work was supported by grants from National Science and Engineering Research Council of Canada, Saskatchewan Agriculture Development Fund, Western Economic Diversification, and Alberta Agriculture Research Institute. P.S.R. was a recipient of a postdoctoral research fellowship from the Health Services Utilization and Research Commission, Saskatoon, Saskatchewan, Canada.

REFERENCES

Akusjarvi, G., and Persson, H. (1981). Control of RNA splicing and termination in the major late transcription unit. *Nature* 292, 420–426.

Anderson, C. W. (1990). Characterization of the adenovirus serotype 2 proteinase polypeptide. *Virology* 177, 259–272.

Anderson, K. P., and Klessig, D. F. (1984). Altered mRNA splicing in monkey cells abortively infected with human adenovirus may be responsible for inefficient synthesis of the virion fiber polypeptide. *Proc. Natl. Acad. Sci. USA* 81, 4023–4027.

Baxi, M. K., Reddy, P. S., Zakhartchouk, A. N., Idamakanti, N., Pyne, C., Babiuk, L. A., and Tikoo, S. K. (1998). Characterization of bovine adenovirus type 3 early region 2B. *Virus Genes* 16, 1–4.

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments of the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* 74, 3171–3175.

Binger, M. H., and Flint, S. J. (1984). Accumulation of early and intermediate mRNA species during subgroup C adenovirus productive infection. *Virology* 136, 387–403.

Boulanger, P., Lemay, P., Blair, G. E., and Russell, W. C. (1979). Characterization of adenovirus protein IX. *J. Gen. Virol.* 44, 783–800.

Caillet-Boudin, M. L. (1989). Complementary peptide sequences in partner proteins of the adenovirus capsid. *J. Mol. Biol.* 208, 195–198.

Cauthen, A. N., and Spindler, K. (1996). Sequence of the mouse adenovirus type-1 DNA encoding the 100-kDa, 33-kDa and DNA-binding proteins. *Gene* 168, 183–187.

Cepko, C. L., and Sharp, P. A. (1983). Analysis of Ad5 hexon and 100K ts mutants using conformation-specific monoclonal antibodies. *Virology* 129, 137–154.

Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V., and Cotton, M. (1996). The complete DNA sequence and genomic organization of the avian adenovirus CELO. *J. Virol.* 70, 2939–2949.

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8.

Chroboczek, J., Bieber, F., and Jacrot, B. (1992). The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology* 186, 280–285.

Chroboczek, J., Ruigrok, R. W. H., and Cusack, S. (1995). Adenovirus fiber. *Curr. Top. Microbiol. Immunol.* 199, 163–200.

Cladaras, C., Bhat, B., and Wold, W. S. M. (1985). Mapping the 5′ and 3′ ends and splice sites of mRNAs from the early E3 transcription unit of adenovirus 5. *Virology* 140, 44–54.

Clarke, M. C., Sharpe, H. B. A., and Derbyshire, J. B. (1967). Some characteristics of three porcine adenoviruses. *Arch. Ges. Virusforsch.* 21, 91–97.

Davison, A. J., Telford, E. A. R., Watson, M. S., McBride, K., and Mautner, V. (1993). The DNA sequence of adenovirus type 40. *J. Mol. Biol.* 234, 1308–1316.

Derbyshire, J. B., Clarke, M. C., and Collins, A. P. (1975). Serological and pathogenicity studies with some unclassified porcine adenoviruses. *J. Comp. Pathol.* 85, 437–443.

Fraser, N. W., Nevins, J. R., Ziff, E., and Darnell, J. E. (1979). The major late adenovirus type-2 transcription unit: Termination is downstream from the last poly (A) site. *J. Mol. Biol.* 129, 643–656.

Gorbalenya, A. E., and Koonin, E. V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Res.* 17, 8413–8440.

Grable, M., and Hearing, P. (1992). Cis and trans requirements for the selective packaging of adenoviral DNA. *J. Virol.* 66, 723–731.

Graham, F. L., and Prevec, L. (1991). Manipulation of adenovirus vectors. *In* "Methods in Molecular Biology, Vol. 7: Gene Transfer and Expression Protocols" (E. J. Murray, Ed.) Humana Press, Clifton, NJ.

Hasson, T. B., Ornelles, D. A., and Shenk, T. (1992). Adenovirus L1 52- and 55-kilodalton proteins are present within assembling virions and colocalize with nuclear structures distinct from replication centers. *J. Virol.* 66, 6133–6142.

Hayes, B. W., Telling, G. C., Myat, M. M., Williams, J. F., and Flint, S. J. (1990). The adenovirus L4 100-kilodalton protein is necessary for efficient translation of viral late mRNA species. *J. Virol.* 64, 2732–2742.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted break points for DNA sequencing. *Gene* 28, 351–359.

Hirahara, T., Yasuhara, H., Matsui, O., Yamanaka, M., Tanaka, M., Fukuyama, S., Izumida, A., Yoshiki, K., Kodama, K., Nakai, M., and Sasaki, N. (1990). Isolation of porcine adenovirus from the respiratory tract of pigs in Japan. *Jpn. J. Vet. Sci.* 52, 407–409.

Hsieh, J. C., Yoo, S. K., and Ito, J. (1990). An essential arginine residue for initiation of protein-primed DNA replication. *Proc. Natl. Acad. Sci. USA* 87, 8665–8669.

Hynes, R. O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* 69, 11–25.

Idamakanti, N. (1998). Molecular characterization of early region-3 of bovine adenovirus-3. Masters thesis. University of Saskatchewan, Saskatoon, Canada.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S. T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* 14, 116–127.

Joung, I., and Engler, J. A. (1992). Mutations in two cysteine-histidine-rich clusters in adenovirus type 2 DNA polymerase affect DNA binding. *J. Virol.* 66, 5788–5796.

Kitchingman, G. R. (1985). Sequence of the DNA binding protein of a human subgroup E adenovirus (type 4): Comparisons with subgroup A (type 12), subgroup B (type 7) and subgroup C (type 5). *Virology* 146, 90–101.

Klessig, D. F., T. Grodzicker, T. (1979). Mutations that allow human Ad2 and Ad5 to express late genes on monkey cells map in the viral gene encoding the 72 K DNA-binding protein. *Cell* 17, 957–966.

Larsson, S., Bellett, A., and Akusjarvi, G. (1986). VA RNAs from avian and human adenoviruses: Dramatic differences in length, sequence, and gene location. *J. Virol.* 58, 600–609.

Linne, T., and Philipson, L. P. (1980). Further characterization of the phosphate moiety of the adenovirus type 2 DNA binding protein from adenovirus type 2 infected cells. *Eur. J. Biochem.* 103, 259–270.

Lu, H., Reach, M. D., Minaya, E., and Young, C. S. H. (1997). The initiator

element of the adenovirus major late promoter has an important role in transcription initiation in vivo. *J. Virol.* 71, 102–109.

Lutz, P., Rosa-Calatrava, M., and Kedinger, C. (1997). The product of the adenovirus intermediate gene IX is a transcriptional activator. *J. Virol.* 71, 5102–5109.

Ma, Y., and Mathews, M. (1996). Structure, function, and evolution of adenovirus-associated RNA: A phylogenetic approach. *J. Virol.* 70, 5083–5099.

Mangel, W. F., McGrath, W. J., Toledo, D. L., and Anderson, C. W. (1993). Viral DNA and a viral peptide can act as cofactors of adenovirus virion proteinase activity. *Nature* 361, 274–275.

Mathias, P., Wicham, T., Moore, M., and Nemerow, G. (1994). Multiple adenovirus serotypes use αv integrins for infection. *J. Virol.* 68, 6811–6814.

McCoy, R. J., Johnson, M. A., and Sheppard, M. (1996a). Genomic location and nucleotide sequence of a porcine adenovirus penton base gene. *Arch. Virol.* 141, 1367–1375.

McCoy, R. J., Johnson, M. A., and Sheppard, M. (1996b). Nucleotide and amino acid sequence analysis of the porcine adenovirus 23K protein. *DNA Seq.* 6, 251–254.

Meissner, J. D., Hirsch, G. N., LaRue, E. A., Fulcher, R. A., and Spindler, K. R. (1997). Completion of the DNA sequence of mouse adenovirus type 1: Sequence of E2B, L1, and L2 (18–51 map units). *Virus Res.* 51, 53–64.

Morin, N., Delsert, C., and Klessig, D. F. (1989a). Mutations that affect phosphorylation of the adenovirus DNA-binding protein alter its ability to enhance its own synthesis. *J. Virol.* 63, 5228–5237.

Morin, N., Delsert, C., and Klessig, D. F. (1989b). Nuclear localization of the adenovirus DNA-binding protein: Requirement for two signals and complementation during viral infection. *Mol. Cell. Biol.* 9, 4372–4380.

Morrison, M. D., Onions, D. E., and Nicolson, L. (1997). Complete DNA sequence of canine adenovirus type 1. *J. Gen. Virol.* 78, 873–878.

Paillard, F. (1997). Advantages of non-human adenoviruses versus human adenoviruses. *Hum. Gene Ther.* 8, 2007–2010.

Pettersson, U. (1984). Structural and nonstructural adenovirus proteins. *In* "The Adenoviruses" (H. S. Ginsberg, Ed.) pp. 205–270. New York, Plenum Press.

Rancourt, C., Tihanyi, K., Bourbonniere, M., and Weber, J. M. (1994). Identification of active site residues of the adenovirus endopeptidase. *Proc. Natl. Acad. Sci. USA* 91, 844–847.

Reddy, P. S., Idamakanti, N., Derbyshire, J. B., and Nagy, E. (1996). Porcine adenoviruses 1, 2 and 3 have short and simple early E-3 regions. *Virus Res.* 43, 99–109.

Reddy, P. S., Idamakanti, N., Derbyshire, J. B., and Nagy, E. (1997). Characterization of the early region 4 of porcine adenovirus type 3. *Virus Genes* 15, 87–90.

Reddy, P. S., Idamakanti, N., Song, J. Y., Lee, J. B., Hyun, B. H., Park, J. H., Cha, S. H., Tikoo, S. K., and Babiuk, L. A. (1998a). Sequence and transcription map analysis of early region-1 of porcine adenovirus type-3. *Virus Res.*, in press.

Reddy, P. S., Idamakanti, N., Zakhartchouk, A. N., Baxi, M. K., Lee, J. B., Pyne, C., Babiuk, L. A., and Tikoo, S. K. (1998b). Nucleotide sequence, genome organization, and transcription map of bovine adenovirus type 3. *J. Virol.* 72, 1394–1402.

Reddy, P. S., Nagy, E., and Derbyshire, J. B. (1993). Restriction endonuclease analysis and molecular cloning of porcine adenovirus type 3. *Intervirology* 36, 161–168.

Reddy, P. S., Nagy, E., and Derbyshire, J. B. (1995a). Sequence analysis of putative pVIII, E3 and fiber regions of porcine adenovirus type 3. *Virus Res.* 36, 97–106.

Reddy, P. S., Tuboly, T., Dennis, J. R., Derbyshire, J. B., and Nagy, E. (1995b). Comparison of the inverted terminal repetition sequences from five porcine adenovirus serotypes. *Virology* 212, 237–239.

Riley, D., and Flint, S. J. (1993). RNA-binding properties of a translational activator, the adenovirus L4 100-kilodalton protein. *J. Virol.* 67, 3586–3595.

Roberts, R. J., Akusjarvi, G., Allestrom, P., Gelinas, R. E., Gingeras, T. R., and Sciaky, D. (1986). A consensus sequence for the adenovirus-2 genome. *In* "Adenovirus DNA: The Viral Genome and Its Expression" (W. Doerfler, Ed.) pp. 1–51. Nijhoff, Boston.

Rohn, K., Prusas, C., Monreal, G., and Hess, H. (1997). Identification and characterization of penton base and pVIII protein of egg drop syndrome virus. *Virus Res.* 47, 59–65.

Russell, W. C., and Kemp, G. D. (1995). Role of adenovirus structural components in the regulation of adenovirus infection. *Curr. Top. Microbiol. Immunol.* 199, 81–88.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.

Shu, L., Pettit, S. C., and Engler, J. A. (1988). The precise structure and coding capacity of mRNAs from early region E2B of human adenovirus serotype 2. *Virology* 165, 348–356.

Smart, J. E., and Stillman, B. W. (1982). Adenovirus terminal protein precursor. Partial amino acid sequence and the site of covalent linkage to virus DNA. *J. Biol. Chem.* 257, 13499–13506.

Sprengel, J., Schmitz, B., Heuss-Neitzel, D., Zock, C., and Doerfler, W. (1994). Nucleotide sequence of human adenovirus type 12 DNA: Comparative functional analysis. *J. Virol.* 68, 379–389.

Stewart, P. L., Fuller, S. D., and Burnett, R. (1993). Difference imaging of adenovirus: Binding the resolution gap between x-ray crystallography and electron microscopy. *EMBO J.* 12, 2589–2599.

Sung, M. T., Lischwe, M. A., Richards, J. C., and Hosokawa, K. (1977). Adenovirus chromation. I. Isolation and characterization of the major core protein VIII and precursor pro-VII. *J. Biol. Chem.* 252, 4981–4989.

Toogood, C. I. A., Crompton, J., and Hay, R. T. (1992). Antipeptide antisera define neutralizing epitopes on the adenovirus hexon. *J. Gen. Virol.* 73, 1429–1435.

Tribouley, C., Lutz, P., Staub, A., and Kedinger, C. (1994). The product of the adenovirus intermediate gene IVa2 is a transcriptional activator of the major late promoter. *J. Virol.* 68, 4450–4457.

Tucker, P. A., Tsernoglou, D., Tucker, A. D., Coenjaerts, E. J., Leenders, H., and Van der Vliet, P. C. (1994). Crystal structure of the adenovirus DNA binding protein reveals a hook-on model for cooperative DNA binding. *EMBO J.* 13, 2994–3002.

Vrati, S., Brookes, D. E., Boyle, D. B., and Both, G. W. (1996a). Nucleotide sequence of ovine adenovirus tripartite leader sequence and homologues of the IVa2, DNA polymerase and terminal proteins. *Gene* 177, 35–41.

Vrati, S., Brookes, D. E., Strike, P., Khatri, A., Boyle, D. B., and Both, G. W. (1996b). Unique genome arrangement of an ovine adenovirus: Identification of new proteins and proteinase cleavage sites. *Virology* 220, 186–199.

Wang, T. S. F., Wong, S. W., and Korn, D. (1989). Human DNA polymerase α: Predicted functional domains and relationships with viral DNA polymerases. *The FASEB J.* 3, 14–21.

Weber, J. M. (1995). Adenovirus endopeptidase and its role in virus infection. *Curr. Top. Microbiol. Immunol.* 199, 227–235.

Weber, J. M., Cai, F., Ramachandran, M., and Burnett, R. M. (1994). Sequence and structural analysis of murine adenovirus type 1 hexon. *J. Gen. Virol.* 75, 141–147.

Webster, A., Hay, R. T., and Kemp, G. (1993). The adenovirus protease is activated by a virus-encoded disulphide-linked peptide. *Cell* 72, 97–104.

Webster, A., Leith, I. R., and Hay, R. T. (1994). Activation of adenovirus-coded protease and processing of the preterminal protein. *J. Virol.* 71, 539–547.

Webster, A., Leith, I. R., Nicholson, J., Hounsell, J., and Hay, R. T. (1997). Role of preterminal protein processing in adenovirus replication. *J. Virol.* 71, 6381–6389.

Wicham, T. J., Mathias, P. A., Cheresh, D. A., and Memerow, G. R. (1993). Integrins $a_v b_3$ and $a_v b_5$ promote adenovirus internalization but not virus attachment. *Cell* 73, 309–319.

Winter, N., and D'Halluin, J. C. (1991). Regulation of the biosynthesis of subgroup C adenovirus protein IVa2. *J. Virol.* 65, 5250–5259.

Wong, S. W., Wahl, A. F., Yuan, P. M., Arai, N., Pearson, B. E., Arai, K., Korn, D., Hunkapiller, M. W., and Wang, T. S. F. (1988). Human polymerase α gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. *EMBO J.* 7, 37–47.

Zhao, L. J., and Padmanabhan, R. (1988). Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. *Cell* 55, 1005–1015.

Zhao, L. J., and Padmanabhan, R. (1991). Three basic regions in adenovirus DNA polymerase interact differentially depending on the protein context to function as bipartite nuclear localization signals. *New Biol.* 3, 1074–1088.

Ziff, E. B., and Evans, R. M. (1978a). Coincidence of the promoter and capped 5′ terminus of RNA from the adenovirus 2 major late transcription unit. *Cell* 155, 1463–1475.

Ziff, E. B., and Fraser, N. (1978b). Adenovirus type 2 late mRNA: Structural evidence for 3′-coterminal species. *J. Virol.* 25, 897–906.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 9, 133–148.